Semantic Web:
Enrichment und Search

von

Gerrit Diederichs

Vortrag im Rahmen des Seminars
Anwendungen 1
des Studienganges

MSc. Informatik

Hochschule für Angewandte
Wissenschaften, Hamburg

Mai 2005

1.Inhaltsverzeichnis

1 Abstract

Der hier zusammengefasste Vortrag ist Teil einer Vortragsserie über das Semantic Web und seiner Technologien im Rahmen des Seminars „Anwendungen 1“ im Sommersemester 2005. Dies ist der dritte Vortrag dieser Serie. Der erste Vortrag von Piotr Wendt mit dem Titel „Semantic Web Services“ hat die Visionen, die mit dem Semantic Web-Ansatz verfolgt werden vorgestellt, und einen Überblick über die grundsätzliche Architektur des Semantic Web in Form eines „Big Picture“ geliefert. Schließlich wurde der Ansatz Web Services unter Zuhilfename von Semantic Web-Technologien zu verbessern vorgestellt. Dieser Ansatz ist unter dem Namen „Semantic Web Services (SWS)“ bekannt. Der zweite Vortrag von Artem Khvat mit dem Titel „Ontologien und Werkzeuge“ hat sich mit den Basistechnologien des Semantic Web beschäftigt. Es wurden Ontologien im klassischen und informationstechnischen Kontext vorgestellt, sowie die Basistechnologien des Semantic Web zur Erstellung solcher Ontologien. Dies sind die Ontologiesprache „Web Ontology Language (OWL)“ und das „Resource Description Framework (RDF)“.
Dieser Vortrag legt seinen Schwerpunkt auf die semantische Anreicherung von Internetressourcen (Enrichment) und die Suche in ontologiebasierten Wissensbasen (Search) im Kontext des Semantic Web. Darüberhinaus wird das Ontologieerstellungstool „Protégé“ das an der Stanford Universität entwickelt wurde vorgestellt. Schließlich der Beitrag für das Projekt „Ferienclub“ erläutert.

2 Motivation

Die Datenflut im Internet wächst täglich. Die Internet Suchmaschine „Google“ indiziert derzeit ca. 8 Milliarden Webseiten. Pro Sekunde werden 17 neue Webseiten generiert (siehe dazu [WLEKLI03] p.13). Die vorliegenden Informationen sind dabei unstrukturiert. Für die Maschinen (Computer) stellen sie lediglich eine Anhäufung von Dokumenten dar, die untereinander verlinkt sind. Die Bedeutung des Inhalts ist bestenfalls in Schlüsselwörtern (HTML Metatag „keywords“) grob umrissen. Die klassische Suche in diesem riesigen, unstrukturierten Datenrepertoire führt immer häufiger zu unbefriedigenden oder sogar unbrauchbaren Ergebnissen. Dieser Sachverhalt wird mit dem Schlagwort „Information Overkill“ beschrieben.
Klassische Suchmaschinen wie oben genanntes Google, Alta Vista o.ä. durchsuchen das Web und indizieren die Seiten mittels einer schlagwort-basierten Volltextsuche. Zusätzlich wird versucht über komplizierte Ranking Funktionen – wie etwa Googles „PageRank“ - eine Gewichtung bezüglich eines Schlagwortes herzuleiten. Im Prinzip bauen diese Suchmaschinen ein gigantisches Stichwortverzeichnis auf. Dabei unterliegen sie folgenden Problemen:

- **Nicht-Einbeziehung von Synonymen**
  Synonyme sind Wort mit ähnlicher oder gleicher Bedeutung. Klassische Suchmaschinen beziehen Synonyme nicht in ihre Suche ein. So können eventuell relevante Informationen nicht angezeigtwerden, da sie garnicht gesucht werden.

- **Ignoranz von Mehrdeutigkeiten (Homonymen)**
  Ein Homonym ist ein gleichlautendes Wort, das aber verschiedene Bedeutungen haben kann. Der Klassiker ist hier das Wort „Java“. Einserseits die Programmiersprache, andererseits auch eine beliebte Urlaubsinsel. Gibt man hier eine Suchanfrage nach dem Wort „Java“ in eine klassische Suchmaschine ein, bekommt man quasi ausschließlich Seiten bezüglich der Programmiersprache Java angezeigt.

- **Ignoranz von Wortformvariationen**
  Worte können auch in verschiedenen Variationen auftreten, beispielsweise „Finanzen“ oder „finanziell“. Klassische Suchmaschinen ignorieren jegliche Wortformvariationen.

- **Nichterkennung sinnverwandter Begriffe**
  Sinnverwandte Begriffe werden von Suchmaschinen ebensowenig wie Synonyme erkannt.

Das Semantic Web stellt einen interessanten Lösungsansatz dar, um die oben genannten Probleme zu lösen und eine effiziente Suche im Web zu realisieren.

3 Lösungsansätze

Um die Suche effizienter zu gestalten, wird maschinenlesbare Semantik hinterlegt. Darauf basierend gibt es zwei grundsätzliche Ansätze die hinterlegte Semantik zu nutzen. Man kann einerseits die Anfrage an sich verfeinern und präzisieren, also auf syntaktischer Ebene eine Verbesserung herbeiführen. Hierfür ist lediglich eine relativ schwache Semantik nötig. Ein Beispiel für den Ansatz die Anfrage basierend auf schwachen semantischen Strukturen syntaktisch anzureichern ist die Diplomarbeit von Andreas Christensen ([CHRISTE05]).
Hier wird eine Topic Map verwendet, um bestimmte Suchbegriffe – in diesem Kontext als Topics bezeichnet – mit zusätzlicher, semantischer Information anzureichern. Suchanfragen werden hinsichtlich bekannter Topics überprüft. Ist ein Topic bekannt, wird die Suchanfrage syntaktisch angereichert und an eine klassische Suchmaschine weitergereicht. Die Vorteile dieser Methode sind die einfache Eingliederung in die bestehende Struktur des heutigen Web und die Möglichkeit klassische Suchmaschinen für die Suche, beispielsweise über die Google API zu nutzen. Allerdings können Suchanfragen bei dieser Methode sehr komplex werden und sind schwer zu formulieren. Die Semantik bei diesem Ansatz ist eher schwach ausgeprägt. So sind beispielsweise keine Inferenzen bezüglich bestimmter Anfragen möglich.
Der zweite Ansatz ist der Semantic Web-Ansatz. Hier wird Domänenwissen in Ontologien modelliert. Anschließend werden Webressourcen basierend auf dem modellierten Wissen annotiert. Die Annotation der Webressourcen kann manuell oder auch weitestgehend automatisiert vorgenommen werden. Inferenzmaschinen ermöglichen es neues Wissen auf Basis des innerhalb der Ontologien modellierten Wissen zu generieren. Folgende Grafik zeichnet das „Big Picture“ des Semantic Web :

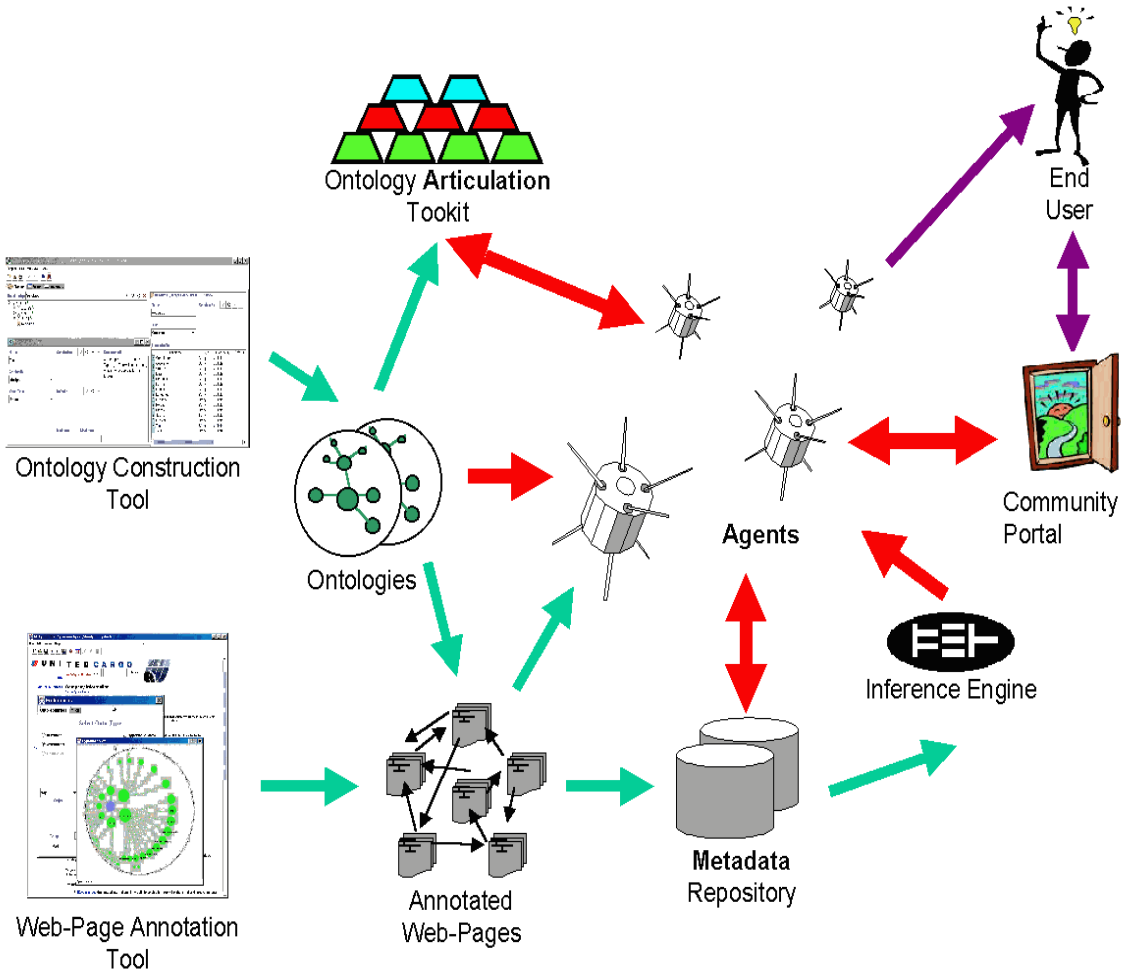

Abbildung 1Big Picture Semantic Web

4 Grundlagen des Semantic Web

Im folgenden werden die Grundlagen des Semantic Web nochmals kurz wiederholt, ausführlich geschah dies bereits im Rahmen des Vortrages von Artem Khvat.

4.1 Resource Description Framework (RDF)

RDF ist eine der Basistechnologien des Semantic Web, obwohl RDF nicht speziell für das Semantic Web entwickelt wurde. RDF stellt ein Modell für Metadaten über Internetressourcen bereit. Das Modell besteht aus Aussagen über Internetressourcen. Eine Aussage besteht hierbei aus drei Teilen:

Subjekt Die Ressource, über die eine Aussage gemacht wird.

Prädikat Eine Eigenschaft der Ressource die hier beschrieben wird.

Objekt Der Wert der Eigenschaft.

Die Aussagen werden also als einfache Sätze in Form von Subjekt-Prädikat-Objekt-Tripeln aufgebaut. Ein Beispiel für eine Aussage ist

„Der Autor von http://dietlweiss.de/ ist Tobias Dietl“

Diese Aussage lässt sich wie folgt einem RDF-Triple zuordnen:

<http://dietlweiss.de/> → Subjekt
<http://terms.org/author> → Prädikat
<http://persons.org/TobiasDietl> → Objekt

Die einzelnen Elemente der Aussage werden dabei i.d.R. durch URLs beschrieben. Neben der Triple-Notation für RDF gibt es zwei weitere Notationen, eine graphische Notation mittels eines gerichteten Graphen und eine XML-Notation. Die in RDF notierten Aussagen über Internetressourcen bilden die Instanzen eines Wissensmodells.

4.2 Standard Ontologiesprachen

Ontologiesprachen werden zur Modellierung des eigentlichen Wissensmodell verwendet. Sie bringen die Semantik in das Modell. Darüberhinaus werden sie verwendet, um verschiedene Ontologien zu mappen, also semantische Überdeckungen mehrerer Ontologien zu kennzeichnen und sie so vergleichbar zu machen. Ontologiesprachen beschreiben Ontologien mittels Klassen, deren Eigenschaften und den Beziehungen zwischen den Klassen. Eine Instanz eines Wissensmodells wird über *<rdf:type>* erzeugt. Die Standardontologiesprachen des Semantic Web sind *RDF Schema* (RDFS) und *Web Ontology Language* (OWL). OWL wiederum ist in drei Sprachen unterteilt *OWL Lite, OWL Description Logic* und *OWL Full.* Bezüglich der Syntax und Ausdrucksmächtigkeit der

Sprachen gilt:

RDFS < OWL Lite < OWL DL < OWL Full

wobei „<“ = syntaktisch und semantisch enthalten

Die Ontologiesprachen bilden zusammen mit RDF die Grundlage für das Semantic Web. Die Ontologiesprachen modellieren das Wissen und ermöglichen Inferenz. RDF-Triples repräsentieren die Instanzen einer Ontologie.

5 Semantic Enrichment

Die semantische Anreicherung von Webressourcen kann generell auf zwei unterschiedliche Arten geschehen, manuell oder weitestgehend automatisiert. In beiden Fällen müssen Ontologien erstellt werden und die anzureichernden Ressourcen müssen für eine korrekte Annotierung klassifiziert werden.

5.1 Manuelle Klassifikation

Bei der manuellen Anreicherung werden sowohl die Ontologien, als auch die Annotationen durch menschliche Benutzer vorgenommen. Die Ontologien werden von Experten für das zu modellierende Wissensgebiet erstellt. Die erste Möglichkeit ist die Annotationen von wenigen Experten in einer zentralen Knowledgebase zu pflegen („Webmaster-Prinzip“). Bei der zweiten Möglichkeit werden die Annotationen von einer Community erstellt und über Annotationsserver zur Verfügung gestellt (siehe hierzu auch die Studienarbeit „Untersuchung von Annotationen und Ontologien im Semantic-Web“ von Nico Richters [RICHTE03]). Das manuelle Enrichment stößt bei großen zu annotierenden Datenmengen wie sie im Web vorliegen an seine Grenzen und ist hierfür unzureichend. Um große Datenmengen zu annotieren muss man das Verfahren weitestgehend automatisieren.

5.2 Automatische Klassifikation

Dieses Kapitel basiert weitestgehend auf der Diplomarbeit von Robert Hoffmann mit dem Titel „Entwicklung einer benutzerunterstützten automatisierten Klassifikation von Web - Dokumenten“ [HOFFMA02].

Bei der automatischen Klassifikation wird das Enrichment der Internetressourcen weitestgehend automatisiert. Normalerweise wird hier eine erste grobe Klassifikation manuell vorgegeben. Basierend auf dieser Anfangsklassifikation werden die Dokumente dann automatisch mittels Textanalyse klassifiziert. Gegebenenfalls wird die Klassifikation selber verfeinert. Das Vorgehen lässt sich grob in zwei Phasen einteilen:

Die Lernphase und die Anwendungsphase.
In der Lernphase wird zunächst ein Set von - für den zu klassifizierenden Themenbereich typischen - Trainingsdokumenten erstellt. Mit verschiedenen Methoden der Textanalyse werden aus diesen Dokumenten verschiedene Attribute extrahiert, die später als Klassifikatoren dienen. In der Lernphase wird so ein erstes Klassifikationsmodell erstellt. In der Anwendungsphase werden dann neue Dokumente anhand des in der Trainingsphase erstellten Modells klassifiziert. Gegebenenfalls wird die Klassifikation durch eine erneute Lernphase verfeinert. Folgende Grafik zeigt den Ablauf der beiden Phasen:

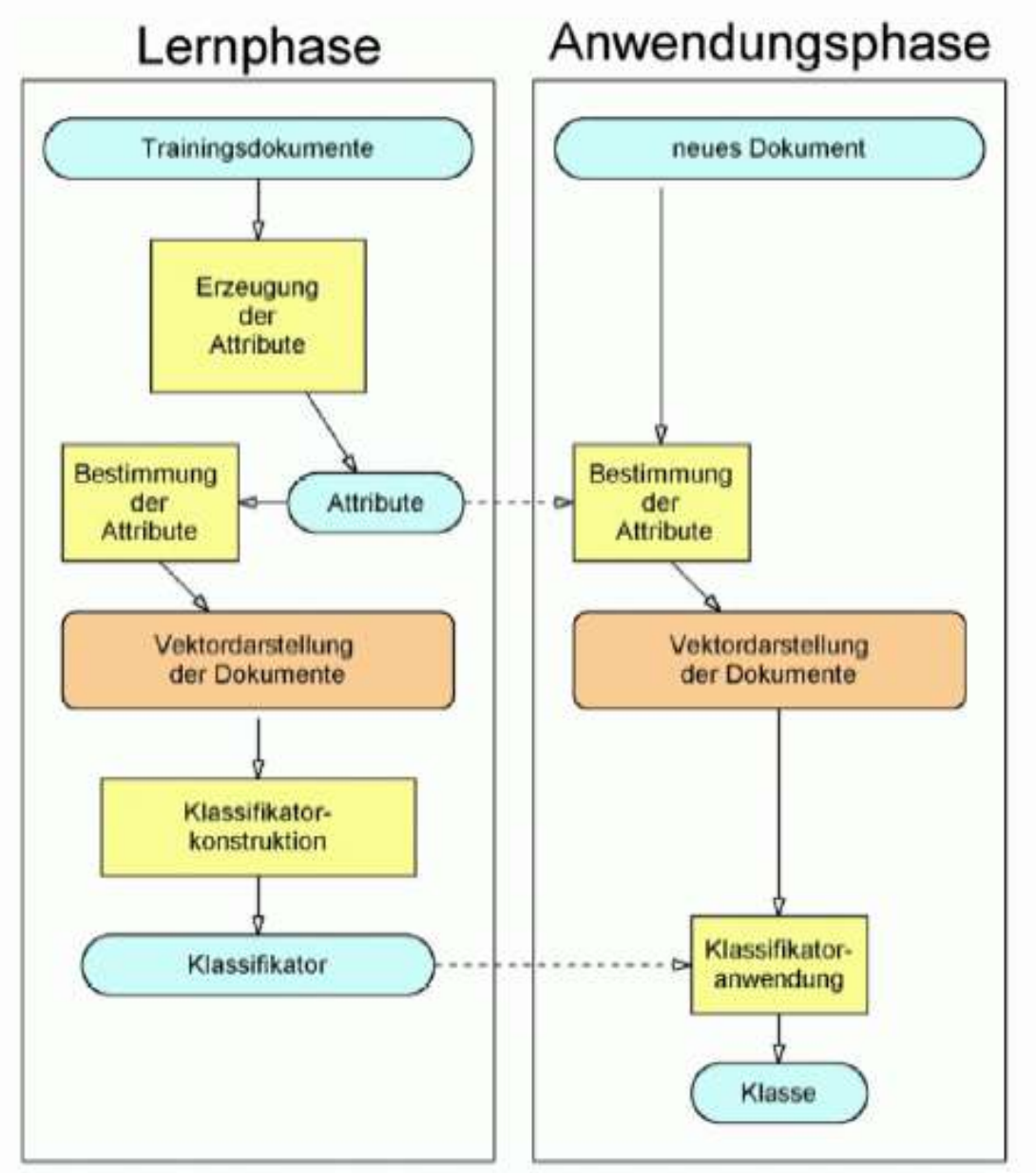

Abbildung 2Die zwei Phasen der Klassifikation

5.2.1 Verfahren der Textanalyse

Die Textanalyse kann im Wesentlichen auf drei Arten erfolgen: Auf linguistischer Basis, auf statistischer Basis oder aber mit Hilfe von begriffsorientierten Verfahren. Die drei Analysearten werden im Folgenden vorgestellt.

Linguistische Analyse

Die linguistische Analyse versucht verschiedene Sprachphänomene wie beispielsweise unterschiedliche Wortformen, Suffixe oder Phrasen zu erkennen. Grundlage sind morphologische, syntaktische sowie semantische Verfahren. Die morphologische Analyse dient dazu nicht sinntragende Worte zu entfernen und somit den Text zu filtern. Zu den morphologischen Analyseverfahren zählen:

- **Stoppwortelimination**

  Besonders häufig oder besonders selten auftretende Worte werden als nicht sinntragend angesehen und entfernt. Beispiele für die deutsche Sprache sind Worte wie der, die, das, mit, in etc.

- **Wortstammbildung**

  Unter Wortstammbildung versteht man die Ermittlung der grammatischen Grund- oder Stammform durch die Rückführung der konkreten Wortform auf einen Wörterbucheintrag. Es ist das umgekehrte Verfahren zur Flexionsformengenerierung.

- **Flexionsformengenerierung**

  Hierbei wird ein Wörterbuch mit allen grammatikalisch möglichen Formen aller Wörter angelegt. Ziel ist es unter anderem, sehr schnell gleiche Grundformen zu erkennen (z.B. die Verben „ging“, „gegangen“ mit der Grundform „gehen“).

- **Kompositärzerlegung**

  Darunter versteht man die Zerlegung von Mehrwortbegriffen auf ihre Wortgrundformen. Schwierigkeiten durch die deutsche Sprache ergeben sich dadurch, dass der vordere Teil aus dem Nominativ oder Genitiv des Singular oder Plural gebildet werden kann.

- **Derivation**

  Es werden Wörter bzw. Wortklassen mit der selben Grundform zusammengefasst (z.B. Berechnung mit rechnen).

- **Phrasenerkennung**
  Hierbei wird versucht Phrasen z.B. mittels Abgleich mit entsprechenden Listen zu erkennen und zu entfernen.

Die morphologischen Verfahren lassen sich in wörterbuchbasierte Verfahren und regelbasierte Verfahren einteilen. Nachdem der Text durch die morphologischen Verfahren gefiltert wurde wird er syntaktisch auf Satzebene analysiert. Schließlich wird versucht auf Dokumentebene eine semantische Analyse durchzuführen. Hierbei wird versucht sinntragende Zusammenhänge des Dokumentes zu extrahieren. Rein linguistische Verfahren sind bei der Analyse natürlicher Sprache zu aufwendig. Sie werden aber häufig in Kombination mit statistischen Verfahren angewandt, um diese zu unterstützen.

Statistische Analyse

Die statistische Analyse beschäftigt sich hauptsächlich mit der Vorkommenshäufigkeit von Wörtern bzw. deren statistischer Verteilung. Die Bedeutung der Wörter wird großteils ausgeklammert. Dadurch erst wird das Problem der Dokumentklassifikation einigermaßen (auch mathematisch) handhabbar. Die eingangs erwähnte Lernphase zum Aufbau eines entsprechenden Modells beinhaltet im wesentlichen 5 Schritte:

- **Die Textnormalisierung**
  Der Dokumenttext wird für die weitere Verarbeitung normalisiert. Es werden z.B. unerwünschte Tags entfernt.

- **Die Termgenerierung**
  Für die Klassifikation werden nicht ausschließlich einzelne Wörter betrachtet, sondern Terme generiert, die die Attribute für die Klassifikation bilden. Methoden hierfür sind Multimengen von Wörtern zu bilden (Bag of words) oder auch n-Gramme.

- **Die Attributauswahl**
  Die Terme bzw. Attribute werden gefiltert, um unbraubare Attribute zu entfernen. Dies geschieht häufig mit den bei den linguistischen Verfahren dargestellten morphologischen Verfahren wie Stoppwortelimination o.ä.

- **Die Attributgewichtung**
  Die übriggebliebenen, relevanten Attribute werden gewichtet. Dies geschieht vor allem auf Basis ihrer Vorkommenshäufigkeit. Ein Verfahren für die Attributgewichtung ist das Term Frequency Inverse Document Frequency (TFIDF) Verfahren.

- **Den Lernschritt**
  In diesem letzten Schritt wird nun mit Hilfe der vorher modifizierten Attribute und eines darauf angewendeten Algorithmus, ein Modell erstellt. Ziel ist also das Trainieren eines Klassifikators für eine feste, bekannte Anzahl an Klassen eines

Trainingsdokumentensatzes. Dies ermöglicht in weiterer Folge auch die Klassifikation von unbekannten Dokumenten.

Es gibt verschiedene Algorithmen für die statistische Klassifikation einige bekanntere sind der Rocchio-Algorithmus, der Naive-Bayes Algorithmus (bedingte Wahrscheinlichkeit), oder das K-Nearest-Neighbor Verfahren. Das statistische Analyseverfahren wird für die automatische Klassifikation am häufigsten verwendet.

Begriffsorientierte Verfahren

Begriffsorientierte Verfahren versuchen die Bedeutung von vorgefundenen Wörtern zu abstrahieren und den Inhalt eines Textes zu erfassen.
Die ermittelte Bedeutung wird anschließend mit Hilfe eines festen Vokabulars (beispielsweise eines Thesaurus oder Wörterbuchs) repräsentiert. Durch die Erfassung der Bedeutung eines Textes einigermaßen unabhängig von den tatsächlich vorgefundenen Wörtern, kommt man damit dem menschlichen Klassifikationsverhalten näher. Allerdings kann hier von tatsächlichem Verstehen noch keine Rede sein. Die moderne Sprachwissenschaft geht von der Annahme aus, die Bedeutung von Wörtern könne nur über den Kontext ihres jeweiligen Gebrauchs erschlossen werden. Optimale Analyseverfahren müssten daher auch den Kontext mitberücksichtigen. Diese Forderung wird weder von linguistischen noch von statistischen Verfahren erfüllt. Die Forschung im Bereich der begriffsorientierten Verfahren geht daher in eine wissensbasierte Richtung. Die Modelle aus dem Bereich der künstlichen Intelligenz zeichnen sich durch Einbeziehung von Weltwissen aus und sind zudem in der Lage, Wissensakquisition zu betreiben. Als nachteilig erweist sich allerdings ein sehr hoher Implementierungsaufwand. Die Systeme sind bereits für den Einsatz in kleinen Systemen extrem aufwendig. Dies ist auch der Grund, weshalb begriffsorientierte Verfahren zur Zeit in der Praxis noch keine große Rolle spielen.

5.3 Fazit

Die manuelle Klassifikation eignet sich nur für relativ kleine Mengen zu klassifizierender Dokumente. Geht es um eine große Menge zu klassifizierender Dokumente, wie z.B. beim Web Mining müssen automatische Verfahren angewendet werden. Hier wird meist ein auf Statistik beruhendes Verfahren benutzt. Ein Beispiel für die automatische Klassifikation basierend auf einem statistischen Verfahren ist das Projekt „Web Fountain“ vom Almaden

Research Center von IBM.

6 Semantic Search

Wie bereits in Kapitel 3 „Lösungsansätze“ erwähnt zeichnet sich der Semantic Web-Ansatz vor allem dadurch aus, dass er die Möglichkeit von logischen Schlüssen (Inferenz) auf Basis der zugrundeliegenden Knowledgebase ermöglicht. Die Grundlagen hierfür (RDF und OWL) wurden in Kapitel 4 kurz wiederholt und im Rahmen des Vortrages von Artem Khvat ausführlich vorgestellt. Für eine Einordnung der Logikkomponente ins Semantic Web eignet sich das Bild des Semantic Web Stack:

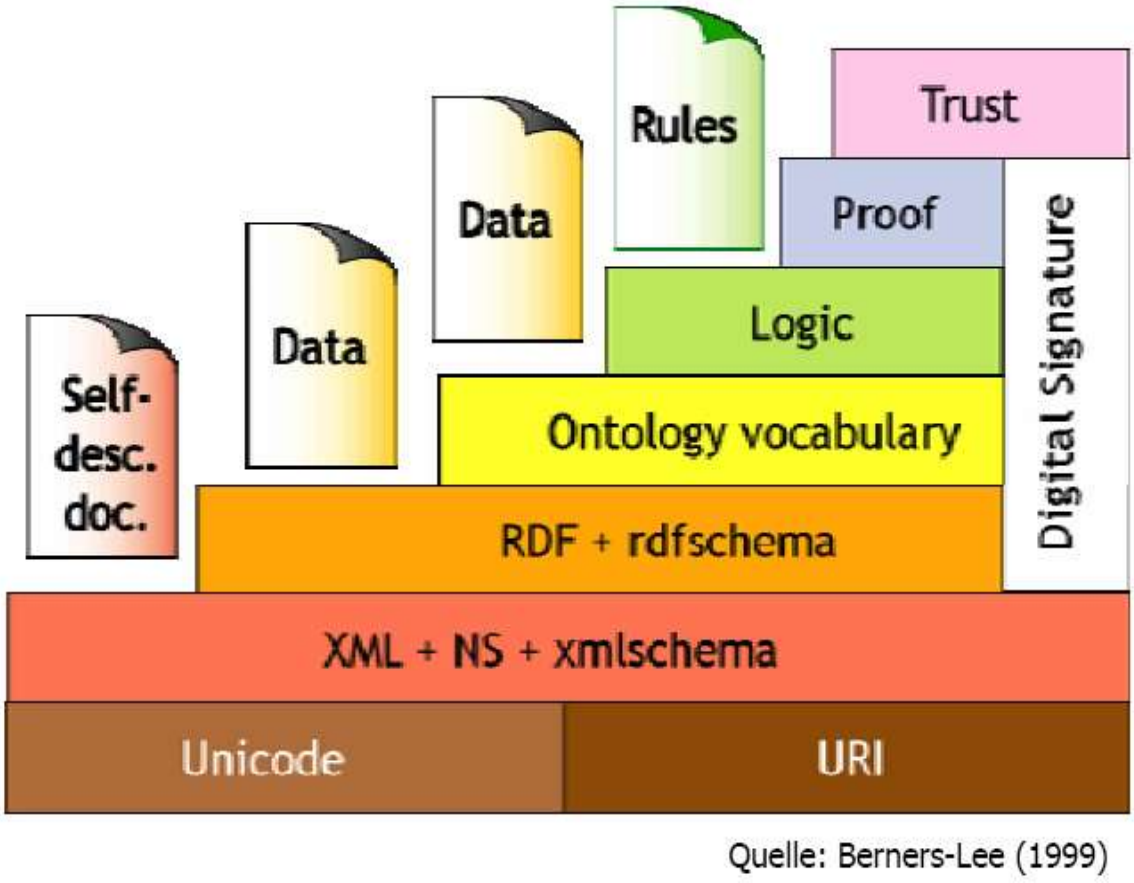

Abbildung 3Semantic Web Stack

Die Erschließung neuen Wissens geschieht nun mittels sogenannter Inferenzmaschinen. Hierbei werden die Inferenzmaschinen hinsichtlich ihrer Mächtigkeit unterschieden. Konkret bedeutet dies sie werden hinsichtlich der Logik, die sie unterstützen unterschieden. Es gibt hier laut [SEMWEB] vier verschiedene Klassen von Inferenzmaschinen:

- **Higher Order Logic Based Inference Engines**
- **Full First Order Logic Based Inference Engines**

- **Description Logic Based Inference Engines**
- **Logic Programming**

Mit Logiken höherer Ordnung (HOL) sind hier Logiken höherer Ordnung als der Prädikatenlogik erster Ordnung gemeint. Mit Full First Order Logic (FOL) ist die Prädikatenlogik erster Ordnung gemeint. Sowohl HOL, als auch FOL sind für bestimmte Fragestellungen nicht entscheidbar. Ich möchte mich an dieser Stelle auf die Beschreibungslogik (Description Logic, DL) beschränken. Um eine Ontologie mittels Beschreibungslogik zu modellieren verwendet man die Sprache *OWL **D**escription **L**ogic* . Neben der Erschließung neuen bzw. abgeleiteten Wissens haben Inferenzmaschinen weitere Aufgaben:

- **Ermittlung der Konsistenz von Ontologien**
  Ist die modellierte Ontologie konsistent ? Auf Klassenebene z.B. die Frage gibt ein Modell in dem die Klasse nicht leer ist ?

- **Ermittlung der Äquivalenz von Ontologien (Mapping)**
  Sind Klassen äquivalent ? Zwei Klassen sind äquivalent, wenn sie in jeder legalen Beschreibung der Welt die gleiche Menge bezeichnen.

- **Klassifikation von Konzepten (Klassen)**

Ist eine Klasse Unterklasse einer anderen ? Dies ist der Fall, wenn für jede legale Beschreibung der Welt eine Klasse eine Teilmenge der anderen Klasse beschreibt.

6.1 Beschreibungslogik

Die Beschreibungslogik ist ein Fragment der Prädikatenlogik. Es gibt dabei nur zwei Arten von Prädikaten: Konzepte (Klassen) und ihre (binären) Beziehungen untereinander (Rollen). Die Beschreibungslogik modelliert somit eine Klassenstruktur. Darüberhinaus gibt es in der Syntax von Beschreibungslogiken keine Variablen. Eine in Beschreibungslogik aufgebaute Wissensbasis (Knowledgebase) besteht aus einer A-Box und einer T-Box. Die T-Box beinhaltet dabei die modellierten Klassen und Rollen, also das Modell. Die A-Box beinhaltet die Instanzen des Modells.
Folgende Grafik zeigt die grundsätzliche Architektur einer mittels Beschreibungslogik aufgebauten Wissensbasis:

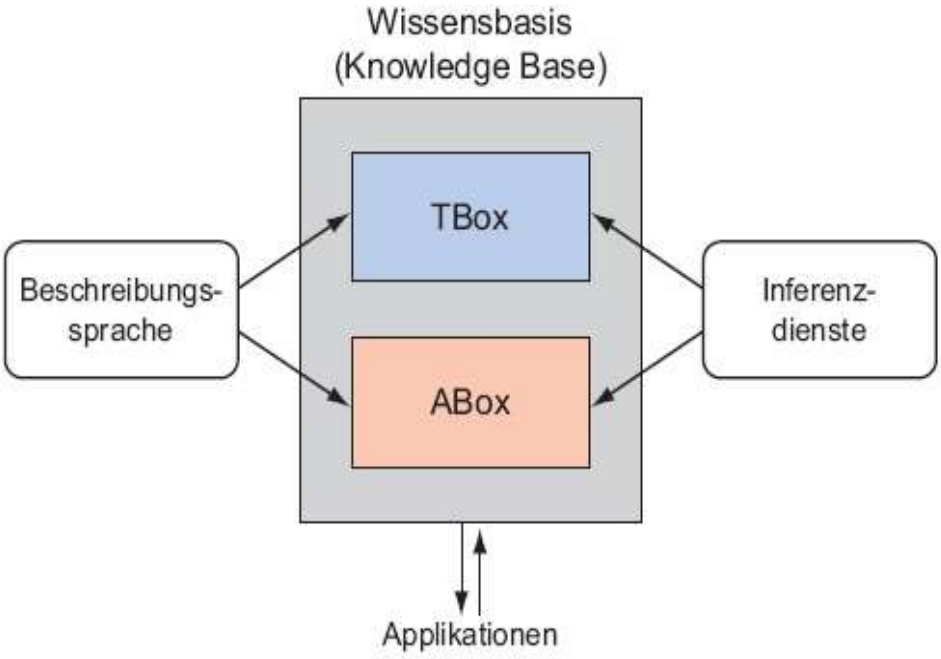

Abbildung 4Architektur eines DL Systems

Die Beschreibungslogik unterstützt die Operationen Disjunktion, Konjunktion und Negation. Ausserdem die Quantoren Allquantor und Existenzquantor.

Ein Beispiel für Beschreibungslogik:

Wenn bekannt ist, dass Rivaner ein Wein ist, aber kein Rotwein und kein Roséwein, kann aufgrund der Vollständigkeit der Zerlegung von Wein in seine Unterkonzepte geschlossen werden, dass Rivaner ein Weisswein ist.

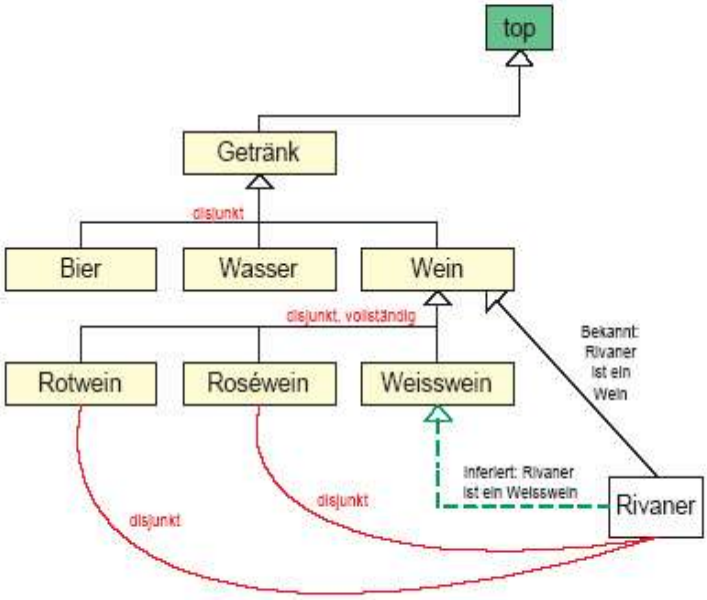

Abbildung 5Beispiel Inferenz mittels DL

6.2 Abfragesprachen des Semantic Web

Die Instanzen einer Wissensbasis im Semantic Web werden in RDF formuliert. Um Anfragen an die Wissensbasis zu stellen Abfragesprachen für RDF verwendet. Diese Abfragesprachen basieren i.d.R. auf einer SQL-ähnlichen Syntax. Exemplarisch soll hier die weit verbreitete Sprache RDF Data Query Language (RDQL) vorgestellt werden.

6.2.1 RDF Data Query Language (RDQL)

RDQL ist syntaktisch an SQL angelehnt berücksichtigt dabei aber die besondere Triple Notation von RDF. Ein RDQL Query setzt sich zusammen aus :

- **SELECT Clause**
  Die Select-Klausel gibt die Namen der Variablen an, die als Ergebnis der Abfrage erwartet werden.
- **FROM Clause**
  Die From-Klausel definiert die URIs der durchsuchten Modelle.
- **WHERE Clause**

Die gesuchten Informationen werden als Triple, dem Grundbaustein von RDF, in einer Where Klausel angegeben.

- **AND Clause**
  Mit AND können Einschränkungen für Variablenwerte angegeben werden.
- **USING Clause**
  URIs sind üblicherweise recht lang. Die USING Klausel definiert Abkürzungen, die in der Abfrage verwendet werden können.

Beispiel eines RDQL Queries:

```
SELECT ?resource, ?familyName
FROM
        <http://example.org/someModel>
WHERE
        (?resource info:age ?age)
        (?resource vCard:N ?y)
        (?y <vCard:Family> ?familyName)
AND ?age >= 24
USING
        info FOR <http://somewhere/peopleInfo#>
        vCard FOR <http://www.w3.org/2001/vcard-rdf/3.0#>
```

6.3 Fazit

Die auf Ontologien basierende Infrastruktur des Semantic Web bietet ein formales Wissensmodell das Inferenzmaschinen nutzen können, um das dargestellte Wissen um implizite Schlussfolgerungen erweitern. Dabei stellen sie eine konsistente und korrekte Wissensbasis sicher.
RDF basierte Abfragesprachen können auf die Instanzen der Wissensbasen zugreifen. Die durch die Semantik ermöglichte Logik bietet eine weitaus mächtigere Alternative als die in Kapitel 3 vorgestellte Anreicherung der Syntax.

7 Angebot für das Projekt

Ich halte das Semantic Web für einen interessanten Ansatz und möchte das Projekt Ferienclub nutzen, um mit dieser Technologie zu arbeiten. Das konkrete Szenario soll zunächst ein Informationsportal für die Clubbesucher sein, das auf einer Semantic Web Infrastruktur aufsetzt. Für das Semantic Web Projekt müssen hierbei einige Dinge geklärt werden. Welche Tools wollen wir für den Aufbau unserer Infrastruktur verwenden ? Gibt es Toplevel oder auch Lowerlevel Ontologien, die für das Projekt verwendet werden können ?
Gegebenenfalls müssen eigene Ontologien modelliert werden, oder auch ähnliche Ontologien gemappt werden. Schließlich muss für das Informationsportal eine benutzerfreundlichere Querysprache als beispielsweise RDQL entwickelt werden.

Mögliche Ausbaustufen wären der Einsatz von personalisierten Agenten, die interessante Angebote für Clubbesucher suchen, oder komplette Ausflüge planen. Auch der Versuch eine teilweise automatische Klassifikation von Webressourcen wäre sehr interessant.

A1.Literaturliste

URLs:
[DIETL02]: http://www11.informatik.tu-muenchen.de /lehre /seminare/seminarSW-SS2002/extension/sprachen.ppt
[GÖTTLI02]: http://www11.informatik.tu-muenchen.de /lehre /seminare/seminarSW-SS2002/extension/logik1.ppt
[FREITA03]: http://www.im.uni-passau.de/lehre/ws0304/DLON/DLON.4in1.pdf
[HOFFMA02]: www.iicm.edu/thesis/rhoff/Hoffmann_DA.pdf
[SCHMUD04] : http://swt-www.informatik.uni-hamburg.de/publications/files/Dipl/Schmude_OntologiebasierteNavigation.pdf
[SEMWEB]: www.semanticweb.org

http://www.w3.org/2001/sw/

Sonstiges:
[CHRIST05]: Andreas Christensen
Diplomarbeit:
Eignung von Topic Maps zur Verbesserung von Suchanfragen
am Beispiel der Studierenden an der HAW im Fachbereich Informatik
[WLEKLI03]: Fabian Wleklinski
Diplomarbeit:
Suche im Semantic Web

Bücher:
Stuckenschmidt, van Harmelen:
Information Sharing on the Semantic Web
ISBN: 3-540-20594-2